Exploratory Analysis - Kitchen Stoves

JJ Espinoza

Sunday, August 23, 2015

The following exploratory analysis examines data on product characteristics of kitchen stoves.

Stove Requirements

Stainless steel

Dimensions of the stove: Depth = 25in, Width = 30in, Height 45in

Key Findings (statistically significant):

- The average price of a stainless steel stove with gas burners is $1,500
- The dimensions of the stove we need is common
- There is a strong correlation between price and weight; likely due to better materials
- The capacity of the stove is also a key driver, specifically the capacity of the oven
- The effectiveness of the stove is important but mainly driven by the heating capacity of the oven

Importing the data.

```
data <- read.csv("C:/Users/ESPIJ090.WDW/Home - Kitchen Stoves/data/Kitchen Stove.csv")
```

We want to limit our search to stainless steel stoves only.

```
stainless <- data[which(data$Color=="Stainless steel" & data$CookingSurface=="Gas: sealed burners"),]
```

Data Visualization and Analysis

What should we expect to pay for a stove?

```
hist(stainless$Price, col = "red", main="Stainless Stove Prices", xlab = "Price($)")
```


```
summary(stainless$Price)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   521.0   914.8  1546.0  1458.0  1762.0  2699.0
```

What are the common dimmensions of stoves in relation to what we are looking for?

We are looking for a stove that is no more than 30in.

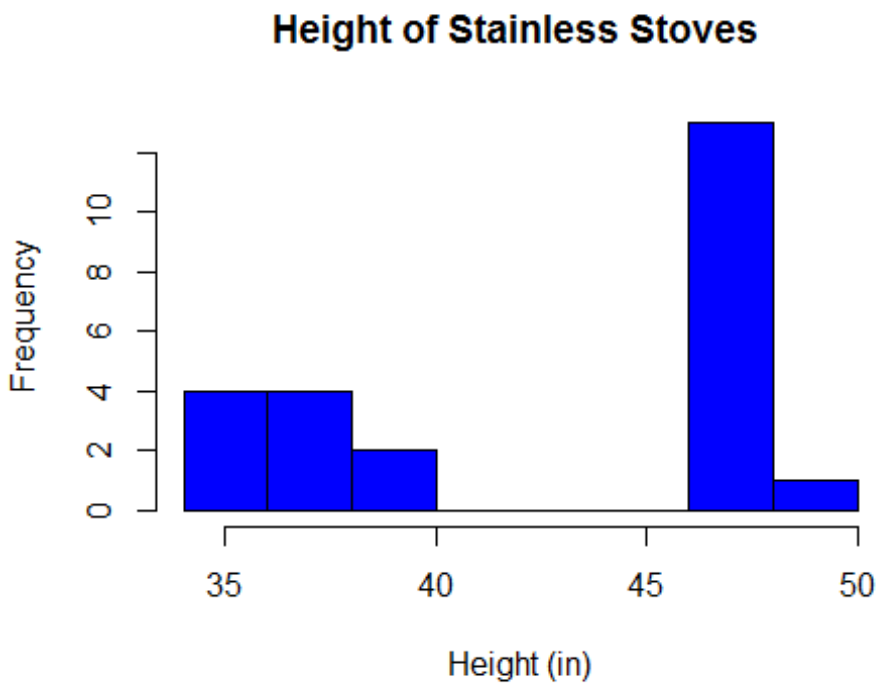
```
hist(stainless$Width, col = "blue", main="Width of Stainless Stoves ", xlab =
"Width (in)")
```


```
summary(stainless$Width)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   29.87   29.90   30.00   30.13   30.00   31.50
```

We are looking for a stove that is close to 45 in height.

```
hist(stainless$Height, col = "blue", main="Height of Stainless Stoves ", xlab
= "Height (in)")
```

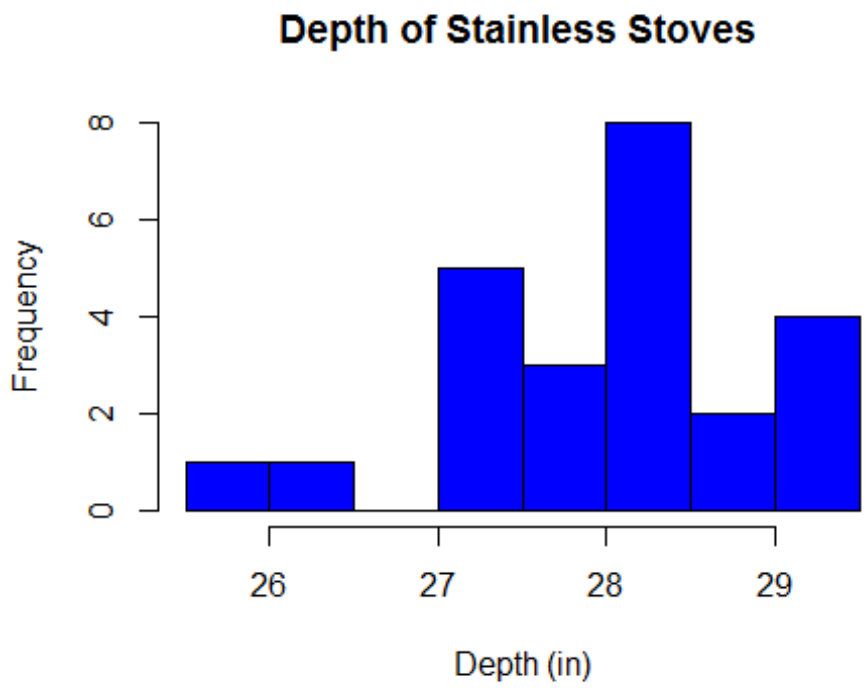



```
summary(stainless$Height)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   35.62   36.62   47.12   43.07   47.50   49.00
```

We are looking for a stove that is close to 25 in depth.

```
hist(stainless$Depth, col = "blue", main="Depth of Stainless Stoves ", xlab =
"Depth (in)")
```




```
summary(stainless$Depth)
```

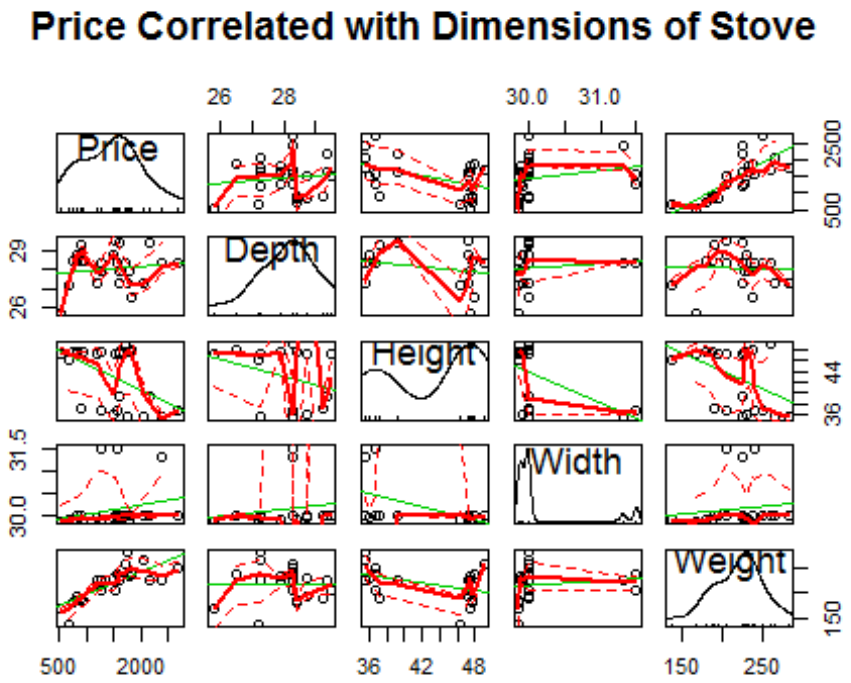
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   25.75   27.25   28.31   28.08   28.49   29.50
```

What are the tradeoffs between stove dimensions and price?

The correlation matrix shows a strong correlation between price and weight.

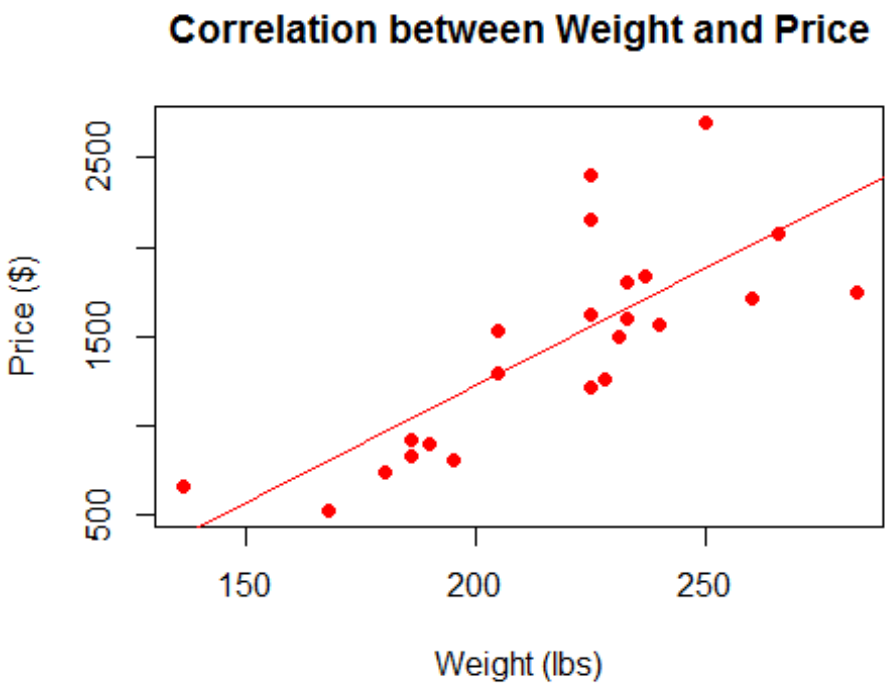
```
library(car)
scatterplot.matrix(~Price+Depth+Height+Width+Weight , data=stainless,
main="Price Correlated with Dimensions of Stove")
```

```
## Warning: 'scatterplot.matrix' is deprecated.
## Use 'scatterplotMatrix' instead.
## See help("Deprecated") and help("car-deprecated").
```



Taking a closer look one can see the clear relationship.

```
plot(y = stainless$Price, x = stainless$Weight, pch = 19, col = "red",main =
"Correlation between Weight and Price", xlab="Weight (lbs)", ylab = "Price
($)")

abline(lm(stainless$Price~stainless$Weight), col="red") # regression line
(y~x)
```



Regression estimate of relationship between price and weight.

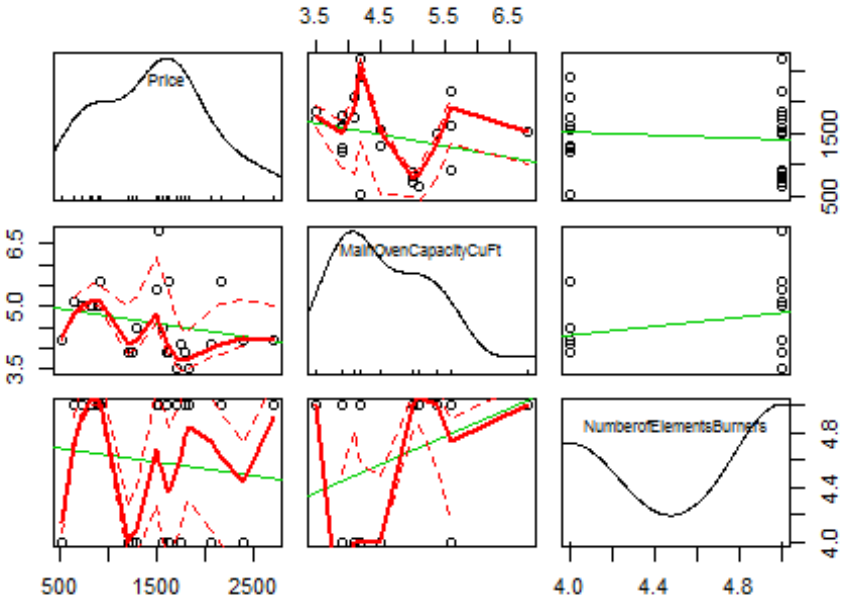
```
summary(lm(Price~Weight, data = stainless))

##
## Call:
## lm(formula = Price ~ Weight, data = stainless)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -560.6 -233.8  -83.3  138.3  852.9
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1411.150    508.159  -2.777    0.011 *
## Weight         13.148      2.303   5.708 9.66e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 367.7 on 22 degrees of freedom
## Multiple R-squared:  0.597,  Adjusted R-squared:  0.5786
## F-statistic: 32.59 on 1 and 22 DF,  p-value: 9.663e-06
```

What other features are correlated with price?

```
scatterplot.matrix(~Price+ MainOvenCapacityCuFt+ NumberofElementsBurners  ,
data=stainless, main="Price Correlated with Dimensions of Stove - Plot1")
```

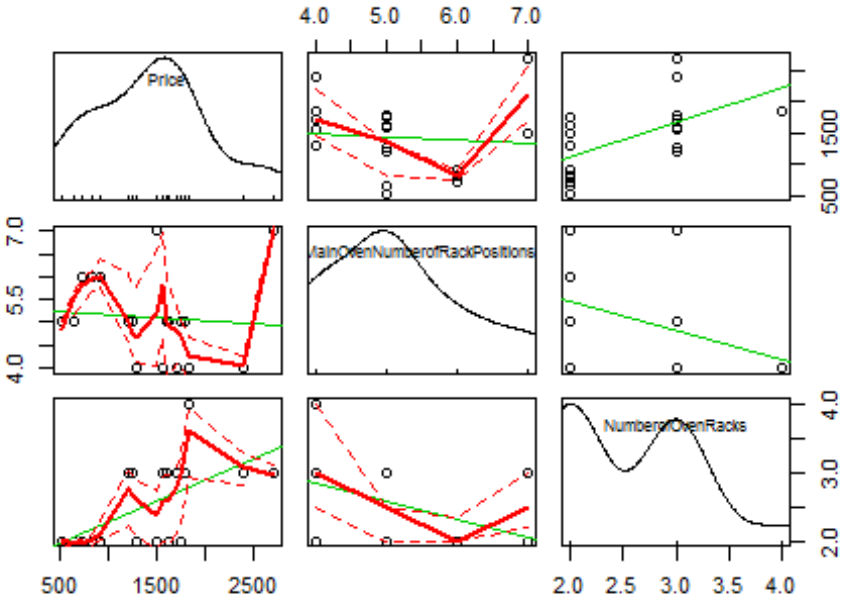
```
## Warning: 'scatterplot.matrix' is deprecated.
## Use 'scatterplotMatrix' instead.
## See help("Deprecated") and help("car-deprecated").

## Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
## spread = spread, : could not fit smooth

## Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
## spread = spread, : could not fit smooth
```

Price Correlated with Dimensions of Stove - Plot1



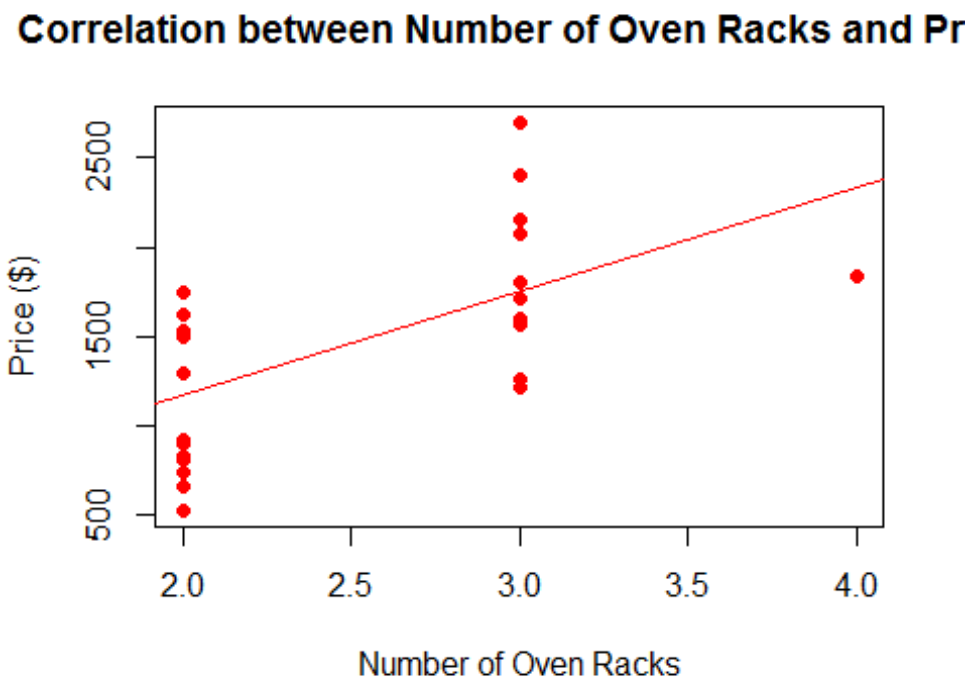

```
scatterplot.matrix(~Price + MainOvenNumberofRackPositions + NumberofOvenRacks
, data=stainless, main="Price Correlated with Dimensions of Stove - Plot2")

## Warning: 'scatterplot.matrix' is deprecated.
## Use 'scatterplotMatrix' instead.
## See help("Deprecated") and help("car-deprecated").

## Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
## spread = spread, : could not fit smooth

## Warning in smoother(x, y, col = col[2], log.x = FALSE, log.y = FALSE,
## spread = spread, : could not fit smooth
```

Price Correlated with Dimensions of Stove - Plot2



Taking a closer look at the correlations betwen Number of Oven Racks and Price we see a negative relationship.

```
plot(y = stainless$Price, x = stainless$NumberofOvenRacks, pch = 19, col =
"red",main = "Correlation between Number of Oven Racks and Price",
xlab="Number of Oven Racks", ylab = "Price ($)")

abline(lm(stainless$Price~stainless$NumberofOvenRacks), col="red") #
regression line (y~x)
```



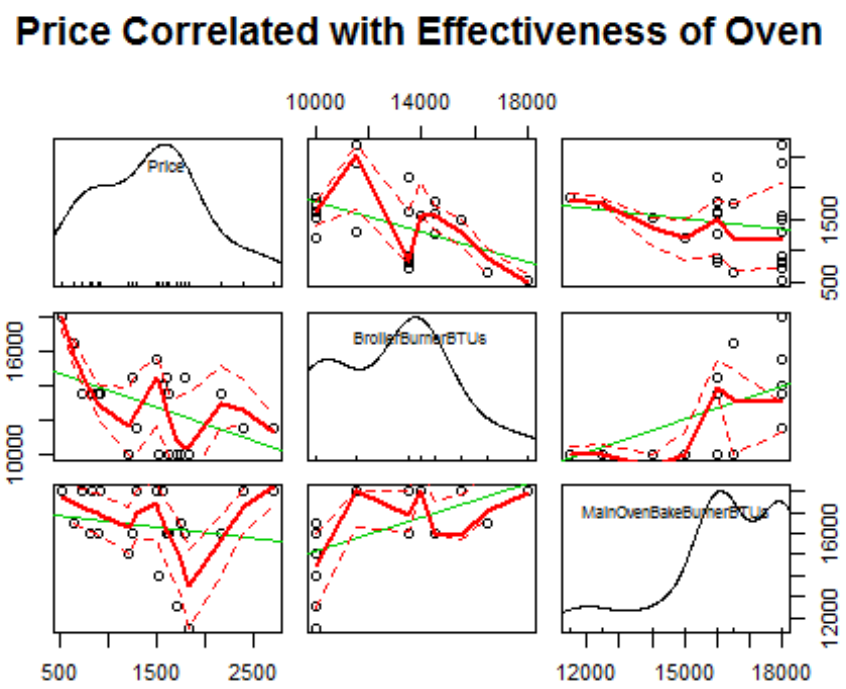
Regression estimate of relationship between price and number of oven racks.

```
summary(lm(Price~NumberofOvenRacks, data = stainless))

##
## Call:
## lm(formula = Price ~ NumberofOvenRacks, data = stainless)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -645.57 -375.82  -94.51  374.32  949.99
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.698    417.854   0.004  0.99679
## NumberofOvenRacks  582.438    162.855   3.576  0.00168 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 460.6 on 22 degrees of freedom
## Multiple R-squared:  0.3676, Adjusted R-squared:  0.3389
## F-statistic: 12.79 on 1 and 22 DF,  p-value: 0.001685
```

How does the effectiveness and efficiency correlate with price?

The scatterplot below shows that the higher the BTUs the lower the price, which seems counterintuitive.

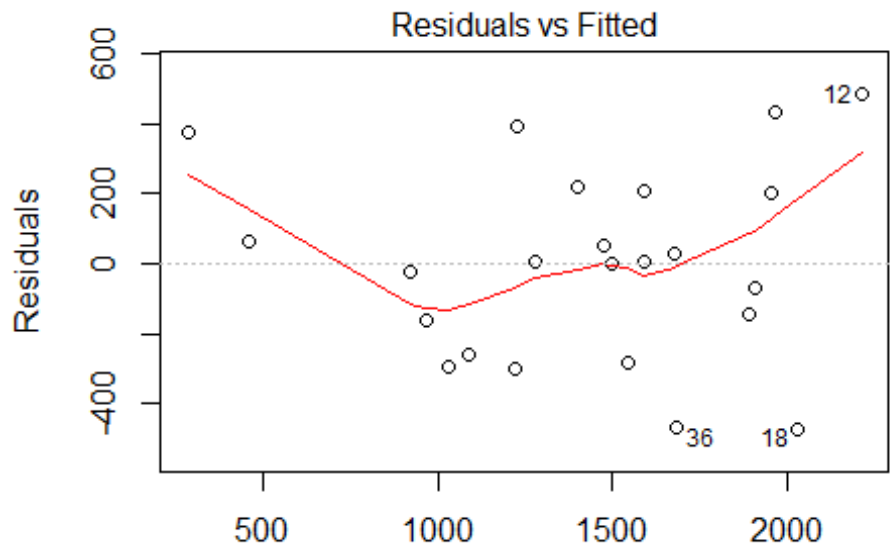
```
scatterplot.matrix(~Price + BroilerBurnerBTUs +  MainOvenBakeBurnerBTUs ,
data=stainless, main="Price Correlated with Effectiveness of Oven")

## Warning: 'scatterplot.matrix' is deprecated.
## Use 'scatterplotMatrix' instead.
## See help("Deprecated") and help("car-deprecated").
```

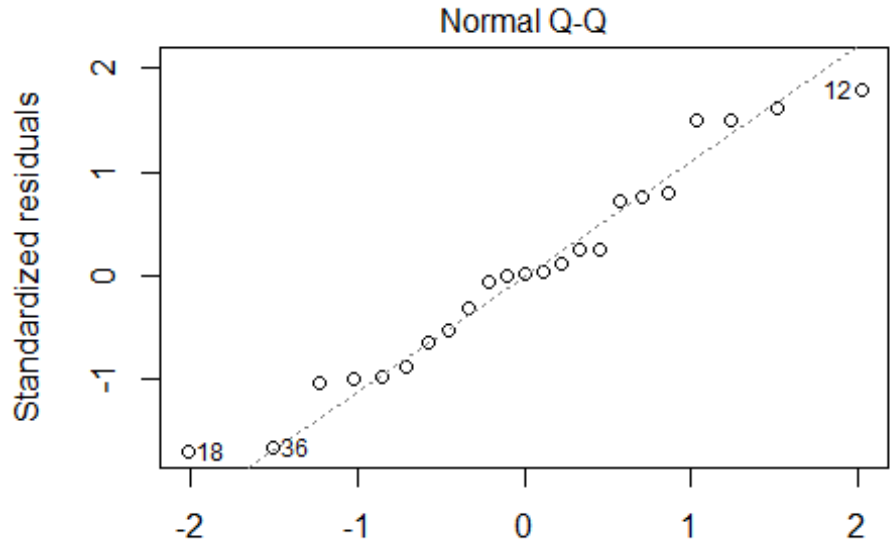


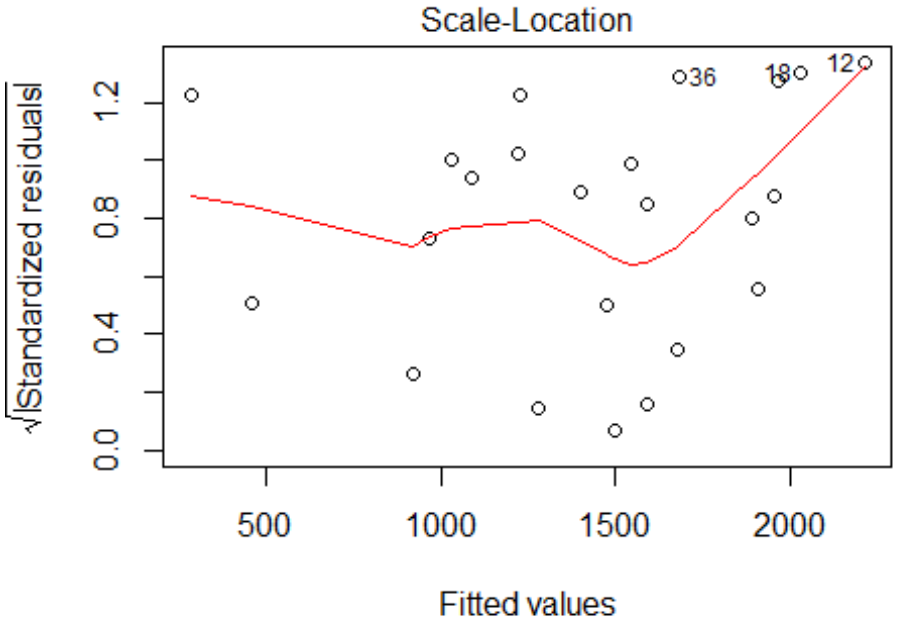
Regression Analysis

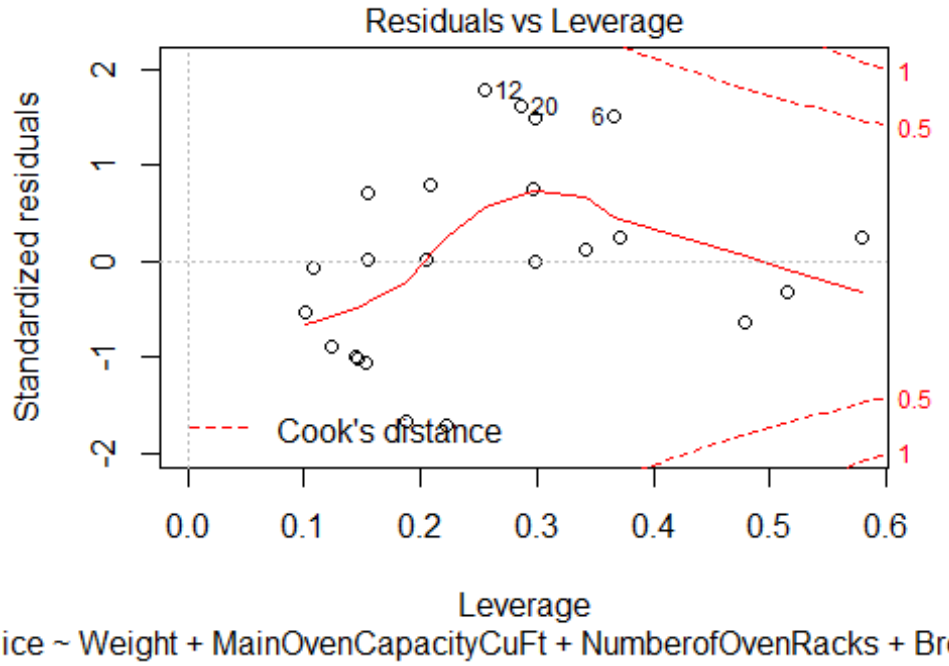
```
model <- lm(Price ~ Weight + MainOvenCapacityCuFt + NumberofOvenRacks +
BroilerBurnerBTUs +  MainOvenBakeBurnerBTUs , data=stainless )

summary(model)

##
## Call:
## lm(formula = Price ~ Weight + MainOvenCapacityCuFt + NumberofOvenRacks +
##     BroilerBurnerBTUs + MainOvenBakeBurnerBTUs, data = stainless)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -470.17 -209.17    6.17  204.00  483.02
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -4.035e+03  1.358e+03  -2.972  0.00855 **
## Weight                  9.836e+00  2.920e+00   3.369  0.00365 **
```

```
## MainOvenCapacityCuFt    2.256e+02  1.020e+02   2.212  0.04097 *
## NumberofOvenRacks       5.592e+02  1.643e+02   3.404  0.00338 **
## BroilerBurnerBTUs      -6.045e-02  4.058e-02  -1.490  0.15459
## MainOvenBakeBurnerBTUs  1.034e-01  4.842e-02   2.136  0.04750 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 313.4 on 17 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.7612, Adjusted R-squared:  0.691
## F-statistic: 10.84 on 5 and 17 DF,  p-value: 8.239e-05

plot(model)
```


Residuals vs Fitted
Residuals
600
200
0
-400
12
36
18
500
1000
1500
2000
Fitted values
ice ~ Weight + MainOvenCapacityCuFt + NumberofOvenRacks + Broi


Normal Q-Q
Standardized residuals
2
1
0
-1
12
18
36
-2
-1
0
1
2
Theoretical Quantiles
ice ~ Weight + MainOvenCapacityCuFt + NumberofOvenRacks + Broi

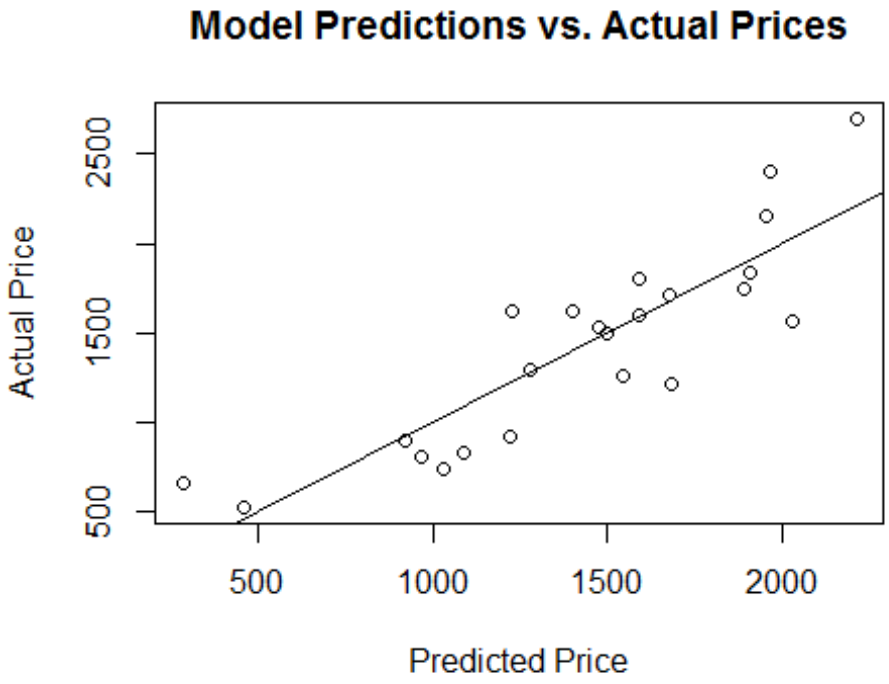




How good does the regression model fit the data?

```
plot(x = predict(model, stainless), y = stainless$Price, main = "Model
Predictions vs. Actual Prices", xlab = "Predicted Price", ylab = "Actual
```

```
Price")
abline(0, 1)
```



The model accuracy is:

```
mean(abs((stainless$Price - predict(model, stainless))/stainless$Price),
na.rm=TRUE)
```

```
## [1] 0.1735026
```

Appendix - Summary Statistics

Summary statistics

```
summary(stainless)
```

```
##         Brand
##  Kenmore   :7
##  GE        :5
##  LG        :4
##  Whirlpool :3
##  Frigidaire:2
##  KitchenAid:2
##  (Other)   :1
##                                                          Name
##  GE 30" Slide-In Gas Range                                 : 2
##  LG 6.1 cu. ft. Double-Oven Gas Range w/EasyClean          : 2
##  Electrolux 30" Gas Slide-In Electric Range w/ Wave-Touch  : 1
```

```
##  Frigidaire 4.2 cu. ft. Freestanding Gas Range               : 1
##  Frigidaire Gallery Gallery 5.8 cu. ft. Double-Oven Gas Range: 1
##  GE 5.0 cu. ft. Gas Range w/ Steam Clean                     : 1
##  (Other)                                                     :16
##          Model.No      Price                    Color        ControlType
##   JGS750SEFSS: 1   Min.   : 521.0   Beige & Bisque : 0             : 5
##  31073       : 1   1st Qu.: 914.8   Black          : 0   Digital   : 3
##  32603       : 1   Median :1546.5   Slate          : 0   Electric  :16
##  32613       : 1   Mean   :1457.8   Stainless steel:24   Mechanical: 0
##  74133       : 1   3rd Qu.:1761.8   White          : 0
##  74233       : 1   Max.   :2699.0
##  (Other)     :18
##                          CookingSurface       OvenCleaningMethod
##  Electric: radiant glass surface: 0     Self-cleaning  :22
##  Gas: open (standard) burners   : 0     Standard clean : 2
##  Gas: sealed burners            :24     Standard clean : 0
##
##
##
##
##      Depth           Height      MainOvenCapacityCuFt     Weight
##  Min.   :25.75   Min.   :35.62   Min.   :3.500        Min.   :136.0
##  1st Qu.:27.25   1st Qu.:36.62   1st Qu.:3.900        1st Qu.:193.8
##  Median :28.31   Median :47.12   Median :4.350        Median :225.0
##  Mean   :28.08   Mean   :43.07   Mean   :4.600        Mean   :218.2
##  3rd Qu.:28.49   3rd Qu.:47.50   3rd Qu.:5.025        3rd Qu.:234.0
##  Max.   :29.50   Max.   :49.00   Max.   :6.800        Max.   :283.0
##
##      Width       BroilerBurnerBTUs MainOvenBakeBurnerBTUs
##  Min.   :29.87   Min.   :10000     Min.   :11500
##  1st Qu.:29.90   1st Qu.:10750     1st Qu.:16000
##  Median :30.00   Median :13500     Median :16250
##  Mean   :30.13   Mean   :12891     Mean   :16354
##  3rd Qu.:30.00   3rd Qu.:14250     3rd Qu.:18000
##  Max.   :31.50   Max.   :18000     Max.   :18000
##                  NA's   :1
##  MainOvenNumberofRackPositions NumberofElementsBurners NumberofOvenRacks
##  Min.   :4.000                 Min.   :4.000           Min.   :2.0
##  1st Qu.:4.250                 1st Qu.:4.000           1st Qu.:2.0
##  Median :5.000                 Median :5.000           Median :2.0
##  Mean   :5.111                 Mean   :4.583           Mean   :2.5
##  3rd Qu.:5.750                 3rd Qu.:5.000           3rd Qu.:3.0
##  Max.   :7.000                 Max.   :5.000           Max.   :4.0
##  NA's   :6
##      NumberofOvens       OvenType
##  Double oven: 8    Convection:14
##  Single oven:16    Standard  :10
##
##
##
```

```
##
##
```